DOCUMENT RESUME

ED 336 088 IR 015 108

AUTHOR Njoo, Melanie; de Jong, Ton
TITLE Support for Learning with Computer Simulations: Giving Hints, Supporting Learning Processes, and Providing Hypotheses.
PUB DATE 91
NOTE 41p.; Paper presented at the Annual Convention of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Students; Computer Assisted Instruction; *Computer Simulation; *Discovery Learning; Discovery Processes; Foreign Countries; Higher Education; *Instructional Design; Instructional Materials; *Intermode Differences; *Learning Processes

ABSTRACT

This paper contains the results of a study on the importance of discovery learning using computer simulations. The purpose of the study was to identify what constitutes discovery learning and to assess the effects of instructional support measures. College students were observed working with an assignment and a computer simulation in the domain of mechanical engineering. Twenty-two learning processes were identified and classified into transformative and regulative categories. In a second study, 91 students received a mixture of directive and non-directive support consisting of an information sheet and fill-in forms. Students were asked to work with the simulation and make use of the forms to support their exploration. By combining the variations in information sheets and fill-in forms, an experimental design was developed containing four experimental groups and one control group. It is concluded that students appeared to be more active in discovery learning when making use of the information sheets and fill-in forms. However, certain discovery processes, such as hypothesis generation, data interpretation, and the drawing of conclusions, are difficult processes and students need assistance to master them. A number of conflicting effects of the instructional support measures were also indicated. (34 references) (Author/DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED336088

Support for Learning With Computer Simulations: Giving Hints, Supporting Learning Processes, and Providing Hypotheses

by

Melanie Njoo and Ton de Jong
Department of Philosophy and Social Sciences
Eindhoven University of Technology
The Netherlands

Paper presented at the Annual Convention
of the American Educational Research Association
Chicago, Illinois, April 3-7, 1991

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Melanie Njoo

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

IR015108

SUPPORT FOR LEARNING WITH COMPUTER SIMULATIONS: GIVING HINTS, SUPPORTING LEARNING PROCESSES, AND PROVIDING HYPOTHESES

Paper presented at the 1991 AERA Annual Meeting, Chicago, April 3-7, 1991.

Melanie Njoo[1]
&
Ton de Jong

Department of Philosophy and Social Sciences
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands
Email wmormn@urc.tue.nl or wmorton@urc.tue.nl

Abstract

Contemporary theories on learning and instruction emphasize the importance of active learners who do not copy information provided to them by a teacher but who discover and structure information and knowledge themselves. This learning approach can be denoted as '*exploratory learning*' or 'scientific discovery learning'. *Computer simulations* create a context that is well fitted for this approach.

The context of the present studies is learning with a simulation program in the domain of control theory, a subdomain of mechanical engineering. Subjects were students at our University of Technology. The studies were conducted as part of their regular curriculum. The aim of the studies was to gain a deeper insight into what constitutes exploratory learning and to assess the effects of a number of *instructional support* measures.

In the first study we made an inventory of exploratory learning processes by observing students working with an assignment and the computer simulation. Further, we analyzed students' thinking-aloud protocols. Twenty-two detailed learning processes were identified. These were classified into two main categories *transformative processes* and *regulative processes*. The study showed that students hardly applied learning processes that we see as characteristic for learning with computer simulations (such as generating a hypothesis). This was also true for a group of students who received additional to their assignment specific hints that were aimed at stimulating these exploratory learning processes.

In the second study, students working with the simulation program received an open assignment that required more free activity. Moreover, students received a mixture of directive and non-directive support that consisted of an *information sheet* and *fill-in forms*. On both sheets and forms, six cells were presented. Each cell for one of the following six learning processes: *identifying variables and parameters*, *hypothesis generation*, *designing an experiment*, *prediction*, *data interpretation*, and *conclusion*. The information sheet provided the subjects with descriptions and definitions of each of these learning processes. There were two variations: sheets with only general information and sheets that additionally contained domain specific (control theory) information. Students were asked to work with the simulation and make use of the fill-in forms to support their exploration. Also, the fill-in forms had two variations: for one all the cells on the set of sheets were empty, for the other we had already filled in the hypotheses cells on the sheets. By combining the variations in information sheets and fill-in forms we had an experimental design with four experimental groups and one control group (total of 91 students). We analyzed the statements on the fill-in forms in a step wise order.

---

[1]The authors like to thank Stephan v. d. Akker for his excellent and laborious work as a domain expert in the studies reported here. We also are thankful to Jos Banens from the Department of Mechanical Engineering for his continuous support and to Wouter van Joolingen for his suggestions and ideas.

First, we conclude that students working with an open assignment, information sheets and fill-in forms appeared to be more active in exploratory learning than was found in Study 1. However, exploratory learning processes are difficult. The data clearly show that the bottle neck in exploratory learning is the application of specific learning processes. Especially the processes of hypothesis generation, data interpretation and the drawing of conclusions are difficult processes. Second, a number of effects of the instructional support measures were found; e.g. we found a higher global activity level (number of forms and cells filled-in) for the groups that were provided with hypotheses on the fill-in forms.

1 Introduction

Learning with computer simulations is considered important in education because it invites learners to engage in *exploratory learning*. Exploratory learning, however, is a complicated study process and it may therefore result in ineffective and inefficient behavior. Therefore, instructional support might be needed in learning with computer simulations. In order to design instructional support it is first necessary to know what is involved in exploratory or 'scientific discovery' learning.

The present paper describes two studies that together cover the following objectives:

a. To gain a deeper understanding of the learning processes that constitute exploratory learning;
b. To examine the effects of different forms of (off-line) *instructional support* for these learning processes. The instructional support consisted of:

- *hints* in a guided assignment;
- *learning processes support* through fill-in forms that divided the study process in separate learning processes, and information sheets that contained either general and/or domain specific information on the learning processes;
- providing learners with *hypotheses* to explore.

The domain in which these studies were conducted was 'control theory'. Subjects were students of Mechanical Engineering at our University of Technology.

1.1 *General background*

Contemporary theories describe the learner as an *active* agent of knowledge acquisition. Knowledge is no longer seen as something that can be copied, but as something that has to be *captured*. *Exploratory* or *scientific discovery* learning is an instructional and learning approach that is well in line with this general notion. In exploratory learning domain information is not offered to the learner in a directly 'consumable' way, but in such a way that the learner has to '*discover*' the information. Exploratory learning is seen as important because of two reasons:

- it is assumed that the domain at hand is learned in a better or 'deeper' way;
- knowledge of the process of exploratory learning itself is seen as an important skill.

The notion of exploratory or (scientific) discovery learning is well known but has recently received renewed attention through the introduction of *computer simulations* as an instructional device. Computer simulations are well fit for discovery learning since they can 'hide' a model that has to be 'discovered' by the learner.

Computer simulations seem to be quite popular in instruction (de Jong, 1990). We feel, however, that quite often computer programs that lack essential characteristics are indicated as being a simulation. According to de Jong (1991b) instructional use of *computer simulations* can be

characterized by four attributes:

- presence of a computational underlying model;
- presence of instructional goals;
- evoking of specific (exploratory) learning processes and,
- providing the possibility for learner activity.

Through these characteristics it is emphasized that learners are active both in their learning processes and in manipulating the underlying model.

Despite their popularity there is too little empirical research on learning with computer simulations and studies that have been conducted quite often do not report straightforward successful results (Rivers & Vockel, 1987; Shute & Glaser, 1990; Swanson, 1990; Veenman & Elshout, 1990). We think that there are two possible grounds for these findings:

- Scientific discovery processes are *too difficult* for learners and because of that they make mistakes and show inefficient and ineffective behavior.
- Learners are *not as active* as we assume and despite the fact that they have exploratory skills they do not use them.

The first reason is the one that is most often reported in the literature. It is found that learners do not work systematically and start floundering. The second one, however, is found as well. Studies show (e.g. Shute & Glaser, 1990; Njoo & de Jong, 1991) that computer simulations do not, simply by nature, invite students to exhibit learning processes that are seen as characteristic for exploratory learning (such as hypotheses generation, predicting, interpreting output etc.). In these studies it was found that learners are rather passive. Of course, this inertness could also be caused by an absence of exploratory skills.

Both types of reason for ineffective or inefficient use of computer simulations as an instructional device need a different remedy. If learning processes are too difficult for learners they need either *training* in these processes or they need *support* at the moment they apply them. If learners are inactive we need instructional measures that *stimulate* them to engage in exploratory learning.

Support to learners can be given in basically two ways: directive and non-directive. *Directive support* guides the learner into a certain direction and, in this way, can take away some of the basic freedom a learner has in exploratory environments. Examples are providing hints and progressive implementation of complexity of models. An extensive survey of directive support measures is given by van Berkum and de Jong (1991). *Non-directive support* does not guide the learner in a specific direction but helps them in activities they intend to perform. An example is a hypotheses scratchpad, on which learners may note down their hypotheses (see de Hoog, de Jong, & de Vries, 1991).

Exploratory learning processes and ways to stimulate and support them are at the heart of our studies and we therefore give them a brief discussion in the following sections.

1.2 *Exploratory learning processes*

Several authors (e.g. Hartley, 1988; Kearsley, 1985; Hille, 1980; Rivers & Vockell, 1987; Lavoie & Good, 1988) indicate that (at a general level) exploratory (or scientific discovery) learning is related to learning approaches such as:

- problem solving
- discovery learning
- inductive learning

These three approaches are similar in that they involve an active, constructive and goal-oriented approach of the learner. In all three approaches learners have to engage in formulating general rules of explicit principles or procedures by themselves. The general idea is that this active, constructive attitude of the learner encourages meaningful incorporation of information into the learner's existing cognitive structure.

We can use studies in these three related fields for inferring what could be important exploratory learning processes.

The relation of *problem solving* with exploratory learning is most clearly described by Cronbach (1977). He sees problem solving as a combination of 'interpretation', 'trial', and 'confirmation'. Two main processes are involved in problem solving: *divergent thinking* (inventing alternative solutions) and *convergent thinking* (reducing the collection of alternative solutions).

Discovery learning is strongly related to exploratory learning. Ausubel, Novak and Hanesian (1978) identify the following processes for discovery learning:

- hypothesis formulation
- hypothesis testing
- determining the strategy of application
- identifying fruitful approaches
- using systematic and economic methods of inquiry.

In the field of *inductive learning* we find the studies by Michalski (1987). Michalski (1987) considers inductive learning to involve *generalizing*, *transforming*, *correcting* and *refining* knowledge representations. Moreover, Michalski distinguishes a number of inductive learning strategies: learning from examples, by experimentation, observation or discovery. These strategies include a variety of processes, such as *creating classifications of given observations*, *discovering relationships and laws governing a given system*, or *forming a theory to explain a given phenomenon*. Michalski also notes that inductive learning mostly involves several concepts and that the learner must be able to *manage* the available time and resources to acquire several concepts at the same time.

Studies that have looked directly at *exploratory* or *scientific discovery* learning are those by Klahr and Dunbar (1988) and Reimann (1989).

Klahr and Dunbar (1988) have proposed a theoretical model that emphasizes the inductive approach within scientific discovery. They propose a general model of scientific reasoning: Scientific Discovery as Dual Search (SDDS). The fundamental assumption is that scientific reasoning requires search in two related problem spaces:

Hypothesis space: The space that represents all the possible hypotheses distinguished by characteristic attributes or frames. Search in the hypothesis space is guided by prior knowledge and experimental results.

Experiment space: The space that represents all the experiments that can be conducted within an informative value range of characteristic attributes. Search in the experiment space may be guided by a current hypothesis and may be used to generate information for the formulation of following hypotheses.

Results of Klahr and Dunbar's studies indicate the existence of 'types' of subjects: 'theorists' and 'experimenters'. Theorists prefer to search the hypothesis space and conduct experiments to test the current hypothesis. Experimenters tend to explore the experiment space, i.e. conducting experiments without an explicit hypothesis. Thus, theorists and experimenters diverge in the way they search for new hypotheses once an initial hypothesis is abandoned: theorists search the hypotheses space for new hypotheses, experimenters search the experiment space to see if they can induce some regularities from experimental outcomes. Experimenters also conduct more experiments than the theorists and experiments are conducted without an explicit hypothesis. The SDDS model presents a scheme of processes and subprocesses ordered into a hierarchy. The three main categories at the top of the hierarchy are:

- Search hypothesis space
- Test hypothesis
- Evaluate evidence

These components are further divided into processes and subprocesses such as '*use prior knowledge*', '*generate experiment*', '*make prediction*', and '*observe*'. The SDDS model can function as a explanation for findings within the field of scientific reasoning such as: failure to seek disconfirmation, tolerating disconfirming evidence and abandoning verified hypotheses.

A similar view is given by Reimann (1989). He studied the behavior of learners in a scientific discovery learning task with a computer simulation. The task required incremental building and testing of hypotheses. Reimann describes the following phases:

- to test and modify hypotheses
- to design an experiment
- to make a prediction
- to evaluate the prediction
- to evaluate/modify the hypotheses.

An important distinction Reimann makes is the one between hypotheses and predictions. In his view a hypothesis is a descriptive generalization over a set of observations and a prediction results from applying a hypothesis to a particular experiment with specific design constraints. A single hypothesis can thus account for several predictions.

However, Reimann's definition of a hypothesis seems limited. In our view it is possible to formulate a hypothesis without any observations when 'observations' are perceived as results from experiments. In fact, this is what the theorists in Klahr and Dunbar's view are doing. Theorists can search the hypotheses space for new hypotheses without actual experimenting. It is likely that Reimann interprets 'observations' as model explorations.

In the first study of the present paper we have observed learners working with a computer simulation and we identified learning processes that they use by analyzing their think aloud protocols. In this analysis we used ideas from the Klahr and Dunbar (1988) and Reimann (1989) study. A third source of inspiration was the work by Lodewijks (1985). Lodewijks makes a broad distinction between learning processes in which information is more or less directly transform into knowledge (*transformative processes*) and processes that are used to control these transformative processes (*regulative processes*).

A final aspect of describing learning processes is the level of detail that is used for the description. What we need is *detailed* knowledge of exploratory learning in order to be able to estimate potential problems in this process and for designing effective support measures. In this respect we will use the

term 'learning process' for indicating *specific* mental actions of learners. As Kirby (1984) indicates, the basic problem is that the concept 'process' can refer to the general category of cognitive or mental functions or to a subset of the cognitive or mental functions. We attempt to avoid the confusion by reserving the term 'learning process' for the specific mental actions (of different levels of detail) and the term 'study process' for the complete knowledge acquisition process.

1.3 *Supporting exploratory learning*

Recently a number of systems that incorporate a simulation as well as additional support have emerged (see e.g. Shute & Glaser, 1990; Towne et al., 1990; White & Frederiksen, 1990). The development of these systems underscore the need for providing help to the learner who is engaged in exploratory learning.

In analyzing this support (see for example van Berkum & de Jong, 1991; de Hoog et al., 1991) we run into a number of seemingly contradictions:

First, it seems obvious that the learners should be *restricted* in what they are allowed to do in order to prevent them from floundering. Examples of such 'restrictions' are providing the learner with increasingly complex models (White & Fredriksen, 1990) or prohibiting that learners enter specified values for some variables (Böcker, Herczeg, & Herczeg, 1989). On the other hand studies show that learners do not by themselves behave as true explorers and show a rather inactive attitude (Njoo & de Jong, 1991). This means that learners need some kind of encouragement by *stimulating* them to perform exploratory actions (e.g., by means of a Socratic dialogue, Finegold & Gorsky, 1989).

Second, a number of studies recommend *directing* the learner in a certain way. This could be done by giving specific feedback on which variable to manipulate next (see for example Towne et al., 1988). On the other hand we see learner control as something paramount to exploratory learning. From this perspective we should design supportive measures that leave as much freedom to the learner as possible. Examples of such measures are all types of scratchpads learners can use. We have termed these types of support *learner instruments* (van Joolingen & de Jong, 1991a).

Third, instructional measures can be imposed upon learners (*obligatory measures*) whereas a different approach could be to leave the control over use of instructional measures in the hands of the learner (*not obligatory measures*).

Adequate support needs to take into account the issues that we raised above and has to find the balance between the extremes. Apart from some exceptions (Shute, 1990) there is little empirical research to the effects of the recommended actions. Moreover, studies that have been performed tend to concentrate on performance measures and not on process measures. The studies that are described in this paper try to shed some light on exploratory learning processes and on the effect of instructional measures.

1.4 *Educational context*

In the studies we report in this paper the domain was a subdomain of mechanical engineering: control theory. Subjects were university students in mechanical engineering. They participated in the experiments as part of their regular curriculum. Subjects worked with a simulation program called PCMatlab (© Mathworks) in a computer lab.

The primary educational goal of the course in control theory is to teach students how to *regulate models of mechanical systems*. Students are taught to regulate the systems by means of a control device, of which they can alter the control law. The regulation influences the relations between the input and output signals of the system. The purpose of the regulation is to obtain an optimal functioning of the system where 'optimal' is described by preferences of the students, some prescribed or mechanical requirements, or the properties of the system (e.g. stable or instable). Figure 1 depicts a global, schematic overview of the domain.

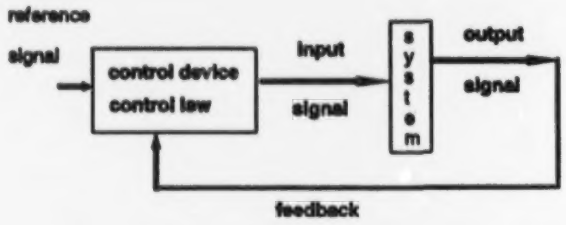


Figure 1 *General model used in control theory*

The *simulation program* students use is *PCMatlab* (© Mathworks). Originally, PCMatlab is not developed for educational purposes but is intended for scientific and engineering numeric calculations and graphics. It can be considered as a type of spreadsheet for complex numeric calculations or as a design tool for constructing and operating simulations of systems. It can, amongst others, be used for linear algebra, matrix calculations and digital signal processing.

PCMatlab offers the student a range of standard functions. In the particular context of control theory, specifications of the system and the control law can be given as differential equations. Once the mechanical system has been given, nothing will be changed in that part of the model. Thereafter, the only input given by the students is the control law or the parameters of the control law. This can result in the following types of output:

- 2- (or 3-) dimensional graphs of variables which describe the relations between input signal and output signal
- numeric data which represent the feedback of the control.

In comparison to traditional instruction in control theory, the use of a simulation program as PCMatlab makes it possible to present the students more complex mechanical systems.

The educational setting in which PCMatlab is used, is a *computer lab*, given parallel to the lectures. The lectures offer the theoretical background of control theory and serve as a foundation for the computer lab. The computer lab consists of one session (3.5 hours) a week for a period of four weeks. During the computer lab, students work in fixed pairs and are free in the choice of their partner. Usually about 10 to 15 pairs are working in a classroom at the same time with two or three tutors available.

Preceding the computer lab, students receive an instruction (3.5 hours). The instruction intends to be an intermediate between the theory and the practical application of the theory in the computer lab. In the instructional session students are informed about the application of rules and methods. The instruction does not involve the operation of PCMatlab.

In Njoo and de Jong (1991) we reported a study on learning with PCMatlab in which we found that students consume a large proportion of their energy for operating the simulation program as such. It was most likely that the excessive high percentage of PCMatlab operation in the previous study was

caused by a deficiency of knowledge of and experience with PCMatlab. Partly as a result of this study the Department of Mechanical Engineering has changed the organization of the computer lab. Three of the major changes are:

- In the first session of the computer lab, PCMatlab operations are explained in a general sense without too much emphasis on control theory.
- The relation between PCMatlab and the domain of control theory is explicitly stressed.
- Students were given the possibility to have experiences with PCMatlab in the first year of their study.

The first session of the computer lab starts with the calculation of simple models. These calculations can still be done by hand. Gradually more complex models are introduced and one needs PCMatlab to calculate these models. Therefore, some PCMatlab operations are presented together with short exercises to practice the PCMatlab syntax. Thus, the importance of the PCMatlab-program is illustrated and the PCMatlab operations are gradually introduced and practiced. The second and third session offer the students the opportunity to increase experiences by employing PCMatlab operations in assignments related to control theory. These assignments are meant as an exercise, offering a model to explore, study questions, and problems to solve. The assignments are not aimed specifically at the stimulation of exploratory behavior: they describe a mechanical system, suggest the students which control laws to use and even which values for the parameters in the control law to explore. After the lab students have to hand in a written report that is discussed with their teacher.

In sum, the educational goals of the computer lab are:

- To acquire knowledge of the principles and procedures of control theory and the regulation of models of systems.
- To acquire knowledge of the principles and procedures of PCMatlab.

Although the second goal is important, the first goal is considered as the primary goal of the computer lab. The subject is taught as a second year course at the Eindhoven University of Technology. Some of the students did not have prior experience with PCMatlab.

2 Study 1: Charting learning processes and the effect of giving hints

The aim of the first study was twofold. First, to test and elaborate a list of exploratory learning processes that we have developed before (Njoo & de Jong, 1991). Second, to evaluate the effects of providing the learners with stimulating hints on the use of specific learning processes. The main reason for introducing the hints was that in our previous study (Njoo & de Jong, 1991) we found that subjects did not behave as 'explorers' and were reluctant to use specific exploratory learning processes. The hints were included in the normal assignment of the students. The study behavior of a group of students who received their original assignment was compared to the behavior of a group who received the same assignment with hints included.

2.1 *Experimental set-up and techniques*

2.1.1 *Subjects*

Subjects in the study were 17 second-year students of mechanical engineering, working in pairs. Subjects participated on voluntary basis. They received no credit points or financial compensation for

their participation.

The 8 pairs (one 'pair' consisted of three students) of subjects were selected at random. To assign the subjects to the two conditions of our study (see experimental setting) we classified them, on the basis of average scores of three prior courses.

2.1.2 *Procedure*

The experiment was conducted as part of the normal curriculum of the students. It was, however, not possible to have the students work in their normal computer lab because we asked them to think aloud and normally 20-30 students work at the same time in the lab. To imitate the natural setting of a computer lab session as closely as possible, subjects were allowed to work in pairs. For experimental control however, pairs of subjects worked together in separate rooms. The tutor was waiting in an other room and was called for whenever the subjects asked for assistance. This was done on purpose, in order to make sure that the verbalization did not influence the tutors actions. Moreover, subjects are now urged to try to solve their problems themselves first, before asking for the tutor.

2.1.3 *Experimental conditions*

Half of the subjects received the original assignment (*unguided condition*), the other half received an altered assignment (*guided condition*). In both condition there were 4 pairs of students.

The alterations within the assignment were aimed at stimulating an exploratory attitude. This was accomplished by giving suggestions for exploration and asking questions about findings. More specifically, suggestions were given to perform the learning processes 'hypothesis generation' and 'test' (see Scheme 1 for a more precise description of these processes). For example, suggestions were given to predict and manipulate:

unguided group	"What is the reaction of the system on a step in u(t)?"
guided group	"What is the relation between de stepresponse and the location of the poles? Make a prediction of the reaction of the system to a step in u(t); verify your prediction. Justify your answers."

The suggestions may seem quite straightforward, but the style of questioning in the original assignment was very direct. For the guided version, we choose to alter the assignment on the aspect of content and not on the aspect of style.

The hypothesis was that, as a result of the altered assignment, the guided group would show significant higher scores in the learning processes of the transformative categories 'hypothesis generation' and 'test' (see Scheme 1).

Subjects were assigned to one of the two conditions on the basis of the average score on three prior courses. The three courses can be considered as introductory courses for the course in control theory. The average scores of subjects in a pair were averaged so that each pair ended with one average end score. We distinguished 'good' and 'poor' pairs. Good pairs had an average end score higher than 7.0 (scores were measured at a 10 point scale where 6 is a pass and 10 is 'excellent'). These groups were evenly assigned to the two conditions so that the unguided and guided group each consisted of two pairs of 'good' and two pairs of 'poor' performers.

2.1.4 *Data collection*

Thinking-aloud protocols were used as the main technique. Furthermore, we had log-files (on-line registrations of subjects input and program output) and the notes subjects made. In order to register the tutor intentions, we asked the tutor to write down his idea of the subjects problem, and the reason why he reacted as he had done. In sum, the data gathered are:

1. thinking aloud protocols
2. tutor notes
3. subjects notes
4. on-line logfiles.

The thinking aloud protocols are the main data source for determining the learning processes. The last three sources were only used to sustain the analysis of the protocols and will not be reviewed in this paper.

The protocols of the 4 pairs of subjects were transcribed and analyzed with the developed analysis scheme. This was done in cooperation with a domain expert to make sure that domain related reflections of the students would be properly interpreted and analyzed. The same method was used in Ferguson-Hessler and de Jong (1990) and showed a very satisfactory interrater reliability.

We used a pair of subjects as the unit of analysis. No discrimination has been made between the subjects within each pair because it was impossible to discriminate between the subjects' line of thought.

2.2 *Results*

The results of this study concern the development of an overview of exploratory learning processes and a comparison of the study behavior in the guided and the unguided experimental groups.

2.2.1 *Overview of exploratory learning processes*

The protocols gathered in this study served as the basis for developing an overview of exploratory learning processes. We also used the overview we developed before (Njoo & de Jong, 1991), and results from literature in creating this overview. Exploratory learning processes were categorized into four categories:

A. Transformative processes
B. Regulative processes
C. Operating the simulation
D. General processes

In the present study the learning processes of the *transformative class* are emphasized. These processes can be characterized as processes of scientific inquiry (e.g. Klahr & Dunbar, 1988) and be classified into four main categories: analysis, hypothesis generation, testing, and evaluation. These processes can be subdivided into more detailed processes (boldface in the table). For example, in the category 'testing', processes for making predictions and data interpretation are involved. Again these processes can be subdivided. *Regulative processes* refer to the executive control of the study process (e.g. planning, monitoring). *Operating the simulation* is concerned with the user interface of the

simulation program. The last class contains processes of a *general nature* (e.g. calculating). At the highest level of detail we identified a total number of 22 processes.

The scheme that follows below present our exploratory learning processes overview.

Scheme 1 *Exploratory learning processes*

Category		Definition
TRANSFORMATIVE PROCESSES		
Analysis		
1.	**Looking for and/or finding of information**	The learner tries to find domain information by searching in text books, additional material, asking a tutor etc.
	Model exploration	Identifying and relating variables and parameters in the model and indicating general properties of the model. This can be done on the basis of prior knowledge, additional material, but of course also at the basis of data from running the simulation.
2.	Model exploration: identifying	Identifying the variables, parameters of the model.
3.	Model exploration: qualitative relation	Making the relations within the model explicit by qualitative statements.
4.	Model exploration: quantitative relation	Making the relations within the model explicit by quantitative statements.
Hypothesis generation		
5.	**Hypothesis generation**	The formulation of a relation between one or more variables (input and output) and parameters in the simulation model. A hypothesis is stated with the intention of testing it.
Testing		Testing involves those activities that are necessary for furnishing data on which the learner expects to be able to accept or refute a new hypothesis, or to create an hypothesis.
6.	**Designing an experiment**	Indicating what will be changed in a simulation model and in which order.
	Making predictions	A prediction states the expectation of a simulation run outcome as the result of designated value attributions to variables.
7.	Qualitative prediction	The expectation is stated in qualitative terms.
8.	Quantitative prediction	The expectation is stated in quantitative terms.
	Learner activities	Changing variable values, etc.
9.	Manipulation of var./internal param.	To handle, control or change the (values of the) variables or internal parameters which describe system properties.
10.	Manipulation of external param.	To handle, control or change the (values of the) external parameters which represent the relation with the environment.
	Data interpretation	Interpreting the data without a direct reference to model relations. The learner can do this in a local manner (noticing specific characteristics of the output, for example: this is an a-symptotic relation) or at a conceptual level by comparing output to other output that is known or to information from other sources.
11.	Local output interpretation	To give one's understanding of the meaning of the output as such.

Category		Definition
12.	Conceptual output interpretation	To give one's understanding of the meaning of the output on a conceptual level. Conceptual interpreting can be done by comparing with other known graphs or data (output from previous assignments or graphs/data familiar from books/lectures).
13.	Supportive processes for interpretation	Performing supportive action for interpretation of the output e.g. changing the range on an axis.
Evaluation		
14.	Evaluating/Judging	Here the actions of the learner and the results thereof are evaluated (e.g., "I shouldn't have done this experiment.").
15.	Generalizing	In generalizing the learner puts his actions and the results thereof in a broader context both as learning processes or as domain information (e.g., "This is an approach I think I can use more often.").
REGULATIVE PROCESSES		
16.	Planning	Planning can be at the level of the complete discovery process, or at the level of one of the phases indicated above. It indicates devising an outline/scheme for what to do.
17.	Verifying	Verifying is checking correctness of actions and results at a conceptual level (e.g., "Did I use the right parameter .... , let's see.").
18.	Monitoring	In monitoring the learner observes and keeps track of his/her own study process (e.g., "So, let's see if we have what we want ...").
PCMATLAB OPERATIONS		
19.	Operations of PCMatlab	Processes for the operation of PCMatlab (user-interface).
GENERAL PROCESSES		
20.	Calculating	Calculating
21.	Making notes	Making notes
22.	Off-task	Off task remarks

2.2.2 *A comparison of exploratory behavior for the guided and unguided groups*

Once developed, the overview of learning processes was used as an analysis scheme for assessing the thinking aloud protocols of the subjects. Utterances of subjects were tape recorded and labelled with the help of the analysis scheme. For reasons of efficiency we had chosen three parts of the assignment for analysis. The selected parts would likely involve most of the explorations of the subjects. The parts of the assignment left out consisted of necessary but rather monotonous activities (preparations for the design of the model and regulation). Unfortunately, most of the subjects did not finish the third and last part of the assignment, which we considered as the most interesting for the analysis.

Table 1 gives an impression of the use that has been made of the different learning processes. Numbers in Table 1 represent mean percentages for the guided and the unguided group separately. The processes as depicted in Table 1 are the processes from the most detailed level from Scheme 1. In Scheme 1 these were indicated with a number.

Table 1
Mean (M) and standard deviation (SD) of the percentages[2] of processes of the analysis scheme for the unguided and guided group

	Groups			
	Unguided ($n=4$)		Guided ($n=4$)	
Processes	*M*	*SD*	*M*	*SD*
Looking for/finding of information	11.8	5.4	12.8	2.2
Model exploration: identifying	0.8	1.7	1.1	0.9
Model exploration (qualitative)	3.4	4.3	3.0	1.9
Model exploration (quantitative)	0	0	0.3	0.7
Hypothesis generation	0	0	0	0
Design an experiment	0	0	0	0
Qualitative prediction	1.6	1.3	0.6	0.7
Quantitative prediction	0.4	0.8	0	0
Manipulation of var./internal parameters	0.0	0	0	0
Manipulation of external parameters	0.0	0	0	0
Local output interpretation	10.1	2.7	6.2	0.8
Conceptual output interpretation	4.9	1.6	2.7	1.9
Supportive processes (int)	1.9	0.7	2.5	2.4
Evaluating/Judging	6.6	2.9	7.9	2.1
Generalizing	0.4	0.8	0.2	0.4
Planning	12.7	5.0	11.2	3.5
Verifying	3.5	1.8	3.5	2.1
Monitoring	16.2	1.4	19.3	4.1
PCMatlab operations	13.5	3.4	9.9	1.3
Calculating	3.9	1.5	6.2	2.9
Making notes	6.1	2.4	7.9	1.8
Off-task	2.4	3.1	4.9	2.7

One of the most important conclusions is that we did not find significant differences between the guided and the unguided groups. Our introduction of hints clearly failed to stimulate true exploratory behavior. This is also evident from Figure 2 in which we present a bargraph that compares the percentages of applied processes from the transformative category (we present the results at the level of detail indicated with boldface in Scheme 1).

The figure shows that exploratory learning processes (such as hypothesis generation, designing an experiment, manipulating of variables) were not applied very often or were (almost) absent. This is in line with what we found in our previous work (Njoo & de Jong, 1991). What we also see is that the subjects need for extra (domain) information was still quite high (about 12%). This may indicate that they were not too well prepared for following the computer lab.

Figure 3 gives the numbers for the regulative processes in a bargraph. The number of planning processes is quite high (especially if we compare this to a non exploratory environment as learning

[2] The percentages are: the absolute number of each learning process related to the total number of learning processes applied by each pair of subjects. The total number of learning processes applied by each pair of subjects differed from 61 to 122, with an average of 84.5.

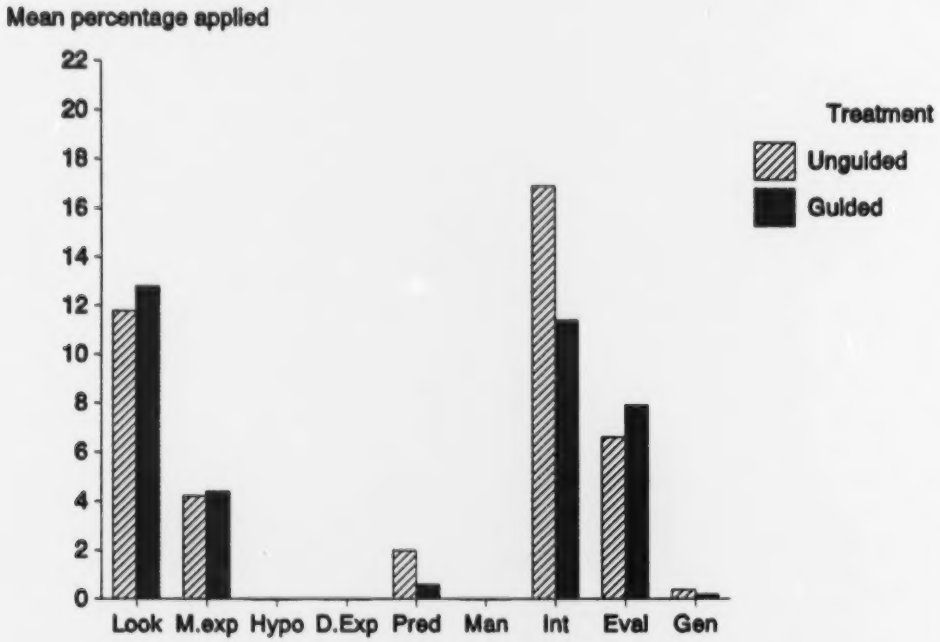


Figure 2 *Percentages of learning processes applied. Comparison between guided and unguided group for transformative processes.*

with text, de Jong & Njoo, 1991). However, as in our previous study (Njoo & de Jong, 1991), subjects tended to use planning on short interval instead of overall strategic planning. Planning related mostly to the general processes such as 'supportive processes for interpretation' and 'calculating'.

A remarkable high percentage was found for the monitoring process (about 17%). However, the monitoring process quite frequently consisted of (re)reading the assignment.

The PCMatlab operations were quite frequently used (11.7%), but not as excessive as in the previous study (Njoo & de Jong, 1991). We have the impression that, as a result of the re-organization of the computer lab, the operation of PCMatlab required less effort than in the previous study.

If we examine the division between the four main categories in Table 1, the balance between the main categories appears to be acceptable. In the previous study subjects used an average of 45% of the processes on the operation of PCMatlab. In the present study an average of approximately 12% appear to be more in congruence with the educational goals of the computer lab. Control theory and the transformative and regulative processes are of more importance than the operation of PCMatlab.

The general conclusion from this study was that providing the learners with hints was not sufficient for stimulating them in applying exploratory learning processes. Therefore, in Study 2 we examined a more drastic instructional measure.

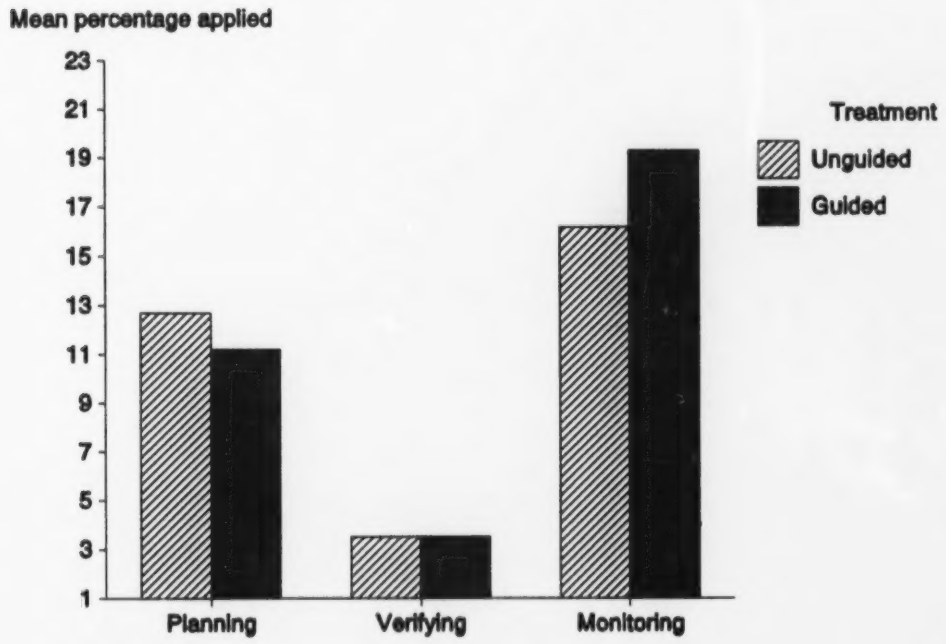


Figure 3 *Percentages of learning processes applied. Comparison between guided and unguided group for regulative processes.*

3 Study 2: Supporting learning processes and the effect of providing hypotheses

Study 1 showed that additional guidance in an existing (closed) assignment did not activate the subjects to generate hypotheses and design experiments. Furthermore, the results of Study 1 showed that the subjects needed domain specific information.

Our general objective in Study 2 was to stimulate exploratory learning processes. Compared to Study 1, the experimental groups in Study 2 were given an assignment which required more initiative and free activity. Since it was an open ended assignment, we expected the subjects to feel less restricted. To stimulate and support the subjects, they were given instructional support for the exploratory learning processes. Support consisted of *fill-in forms* and *information sheets*. The effects of these instructional support measures on quantity and quality of exploratory learning processes was assessed.

3.1 *Experimental set-up and techniques*

3.1.1 *Subjects*

Subjects in Study 2 were 91 second year students of mechanical engineering. Subjects worked in pairs and this resulted in 44 pairs (some 'pairs' consisted of three students). Four experimental groups (*n*

is 7 to 10 pairs of subjects) and a control group of 10 pairs participated voluntary.

3.1.2 *General procedure*

All groups (experimental and control) participated in the experiment (a computer lab session of 3.5 hours) as part of their normal curriculum. During the lab session subjects could ask assistance of a tutor.

The control group followed the simulation lab in a traditional way with a normal assignment and no additional support (for a description see Section 1.4). All subjects of the experimental groups received:

- specifications of a *modelled system* and
- an *open ended assignment* to explore the modelled system with the aim of constructing the optimal regulation for the system. It was stressed to these subjects that this assignment was different from other assignments they had received in the lab (see Section 1.4) and that they were free to explore the system as they wanted. It was also emphasized that their explorations did not necessarily have to involve the optimal regulation but could also involve less optimal regulations. We informed them that the explorations of other (non-optimal) regulations could result in more insight in the system and regulation in general and that this insight could help them to explain their final choice. More specifically the modelled system and the assignment concerned a ship that had to be kept on course.
- *additional support* in the form of an information sheet and fill-in forms. Different variations of the information sheet and fill-in forms were designed. Experimental groups differed in the specific variation they received (see the next section).

All groups made a post-test which consisted of seven multiple-choice questions which tested qualitative insight in the domain.

3.1.3 *Experimental conditions*

In the experimental conditions subjects were provided with a mixture of directive and non-directive instructional support for exploratory learning processes. The support was given off-line (paper and pencil) and it consisted of an *information sheet* and *fill-in forms*. An information sheet was offered to the subjects at the start of the lab session. This sheet contained information on a number of exploratory learning processes. After having read the information sheet subjects were asked to work with the simulation using the fill-in forms. Fill-in forms offered the learners the opportunity to note down their thoughts, actions, or results of the simulation for each of the exploratory learning processes that were explained at the information sheet.

Structure of information sheets and fill-in forms
The structure of both information sheets and fill-in forms was identical: the sheets and forms were divided into six cells labelled VARIABLES & PARAMETERS, HYPOTHESIS, EXPERIMENT, PREDICTION, DATA INTERPRETATION, and CONCLUSION. This subdivision was based on the specific learning processes as identified in Study 1 (see Scheme 1). The size was 16.5 x 11.8 inch (42 x 30 cm). The general structure looked as follows:

VARIABLES & PARAMETERS	HYPOTHESIS	DATA INTERPRETATION
EXPERIMENT	PREDICTION	CONCLUSION

Information sheets
The cells of the information sheet contained information on the six exploratory learning processes. We designed two variations of the information sheet. In one variation we only offered *general information* about the six learning processes. In the other one learners were not only provided with this general information, but additionally with *domain specific information* on the processes.

Two examples will clarify this difference. The examples are from the cell VARIABLES & PARAMETERS. The first example is from the general part of this cell:

> (...) A model is described by variables and parameters. Variables represent the state of the system and can be classified as dependent and independent variables. The independent variables are not influenced by other variables in the model. (...)

The second example is a part of the domain specific information given in this cell:

> (...) The independent variables could be:
> The time t;
> The input signal alpha(t) and it's Laplace transform A(s); (...)

Fill-in forms
The set of fill-in forms came in two variations:

a. forms that had only blank cells with the six labels (essentially as shown above as a general structure) *(free fill-in forms)* and
b. forms that had the cell HYPOTHESIS already filled in (*hypotheses fill-in forms*). Each pair of subjects received nine fill-in forms, so s/he had nine different hypotheses already given (for each of the nine fill-in forms the cell HYPOTHESIS already contained an hypothesis).

Thus, the groups with free fill-in forms had to generate hypotheses for exploration themselves. The

groups with the hypotheses fill-in forms could explore the model with the help of the hypotheses that were provided but they were also free to generate other hypotheses. In order to do the latter, they also received a few free fill-in forms.

The hypotheses that we offered the subjects gave different viewpoints on the model and the different possible regulations for the system. The hypotheses were stated in an affirmative or a negative sense and they could be verified or falsified. Examples of hypotheses that we provided are:

"With a proportional control law you do not have influence on the stability of the system."

"The value of the feedback amplification K has influence on the sub or super critical damping of the system."

Experimental groups
The two main experimental conditions we created were *information sheets* (general or domain specific) and *fill-in forms* (free or hypothesis). Combining these conditions resulted in the following experimental design:

	General information sheet	Domain specific info sheet
Free fill-in forms	n = 7 pairs	n = 8 pairs
Hypotheses fill-in forms	n = 10 pairs	n = 9 pairs

The experimental groups (that were existing computer lab groups) were assigned to the conditions at a random basis.

Procedure in the experimental groups
Subjects were instructed to read the information sheet carefully and to work through the assignment by filling in forms from the set of fill-in forms. There was no compulsory order of cells and subjects were urged to follow the order they preferred. Figure 4 gives a pictorial representation of the experimental simulation.

3.1.4 *Data collection*

For assessing the effects of the different instructional support measures on the exploratory study process we analyzed the quantity and quality of the statements that the subjects noted at the fill-in forms. This analysis was performed in cooperation with a domain expert.

The analysis was performed in a step wise order. This means that we introduced a number of levels for the analysis and at each level a specific aspect of the statements was assessed. Some of the levels are based on the 'learning indicators' of Shute and Glaser (1990).

At the first level we looked at the *global activity* of the subjects. This was assessed by counting the number of forms that were filled in and the number of cells filled in at each form. At the second level we estimated if statements of the subjects were in line with 'definitions' (as given at the information sheet) of specific learning processes. So, for example, at this level it was rated if a statement made in the cell HYPOTHESIS was really an hypothesis. In order to be able to make a correct assessment, a small 'assessment scheme' was created for each of the cells separately and each statement of a subject was rated against this assessment scheme (the assessment scheme can be found in Appendix 1). This second level was termed *learning process validity*. At the third level we looked

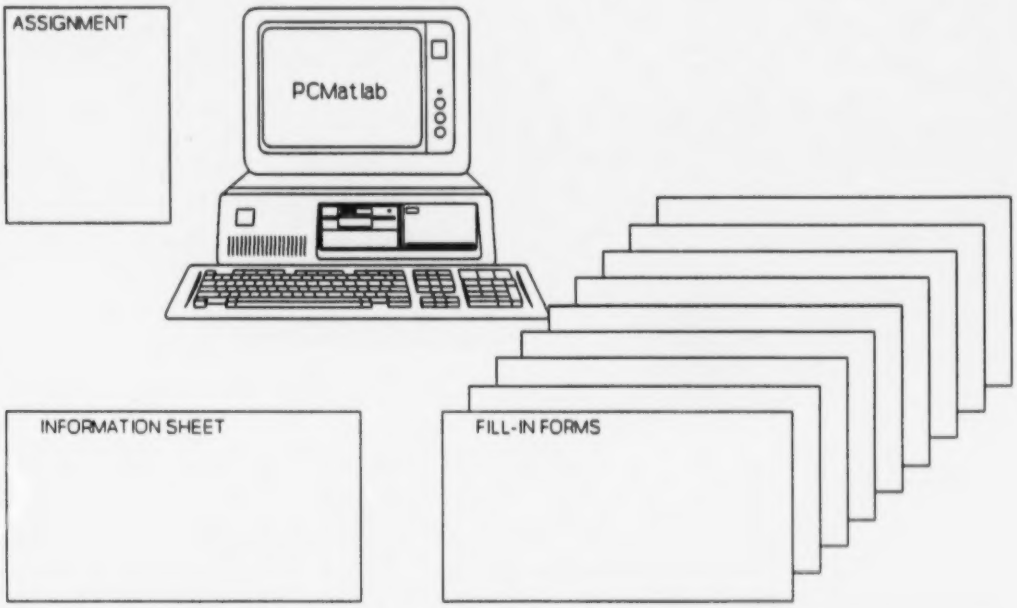


Figure 4 *Subjects received one information sheet, nine fill-in forms and an assignment.*

at the *domain correctness* of the statements. In the first three levels individual cells were assessed, at the fourth level we looked at the relation between cells. *Consistency* stated in how far statements in different cells at one fill-in form were related to each other. Finally, at the fifth level, we analyzed the relations between different fill-in forms for the same pair of students. The *overall strategy* expressed how hypotheses stated at different fill-in forms for one pair of students were 'logically' related. Scheme 2 gives an overview of the different levels of data analysis.

Each level gives a more precise and close analysis of subjects' input. The general idea behind the different levels of analysis was that each level would work as a *'sieve'*; statements that were not valid at a certain level would not be analyzed at a next level. The only exception to this general idea are the 'false' hypotheses (in the sense of 'hypotheses that can not be verified'). Hypotheses, however, are analyzed separately from the other cells for the domain correctness level.

At each level the 'qualitative' assessment of the subjects' statements was summarized into a quantitative score as is also depicted in Scheme 2. The 'sieve' idea also influenced the calculation of these scores. Since the scores on each level are related to the scores on the previous level we have used *relative scores*. For example: scores on the learning process validity level are related to the maximum score the subjects could have achieved, given the number of cells they have used (and which was scored on the previous global activity level). So, the number of forms is the only absolute score. Other scores are related to the scores on the previous level. We took into account that the groups that received the hypotheses fill-in forms were given hypotheses to explore and could not score in the HYPOTHESIS cell.

Scheme 2 *Analysis levels, data and operationalisation*

Level of analysis	*Data*	*Operationalisation*
Global activity assessment of the general activity level of the subjects	• the number of forms • the total number of cells	
Learning process validity assessment of the level in which the statements in each cell answered the general description of the cell (as was given on the information sheets)	• scores per cell • total score	For each cell we determined what aspects should be present. For example: in the cell EXPERIMENT subjects could note down the input variables, the output variables, and the values of the input variables (see Appendix 1 for the other cells). For the six cells on one sheet, a total of 20 points could be scored.
Domain correctness assessment of the domain correctness of the statements	• scores per cell • total score	For the aspects of the learning process validity level we determined the domain correctness. For example: If the value(s) of the input variable(s) was/were given in an experiment, we scored the domain correctness of this choice. In this example, a mistake would be to give a negative value if the input variable could only have positive values. Of course, subjects could state 'false' hypotheses as well. Therefore, the hypotheses were analyzed separately from the rest of the data.
Consistency assessment of the relations between contents of different cells at one fill-in form	• total score	Only learning process valid and domain correct statements were analyzed at this level. We determined at a global level if there was a relation between the statements in the two cells. For example: are the same variables used in the hypothesis and the experiment. The following relations were evaluated: HYPOTHESIS - EXPERIMENT EXPERIMENT - PREDICTION EXPERIMENT - DATA INTERPRETATION PREDICTION - DATA INTERPRETATION DATA INTERPRETATION - CONCLUSION The relation HYPOTHESIS - EXPERIMENT was analyzed separately from the rest of the data. Valid but 'false' hypotheses were analyzed at this level.
Overall strategy assessment of the development of the statements in the same cell through different forms	• total score	The development of the successive hypotheses and the relation between a hypothesis and the conclusion on the previous form.

3.2 *Results*

3.2.1 *Global activity level*

At the first level of analysis the number of forms and the number of cells that were filled in were scored as an indication of the activity level of the subjects. For each pair of subjects the number of cells filled in was related to the maximum number of cells they could have filled in taken into account the number of forms this pair had used. In this way we calculated a percentage of cells filled-in. In

this, we have taken into account that the groups with hypotheses fill-in forms could not fill in the *HYPOTHESIS* cell.

Table 2 shows that overall the subjects used an average of almost 5 fill-in forms. According to the time they had for the assignment, it took an average of about 40 minutes to complete one form. Table 3 shows that almost 85 % of the cells on the forms were used. On each form the subjects used an average of about 4.5 cells (the groups with the hypotheses forms only could fill in five cells).

Table 2
Mean scores (M) and standard deviations (SD) for the number of forms used

	Fill-in forms		
Information sheets	Free	Hypotheses	Total
General			
M	3.7	5.2	4.6
SD	1.3	1.1	1.4
	(*n* = 7)	(*n* = 10)	(*n* = 17)
Domain specific			
M	3.6	5.8	4.8
SD	0.9	1.8	1.8
	(*n* = 8)	(*n* = 9)	(*n* = 17)
Total			
M	3.7	5.5	4.7
SD	1.1	1.5	1.6
	(*n* = 15)	(*n* = 19)	(*n* = 34)

Table 3
Mean scores (M) and standard deviations (SD) of the percentages of cells filled in

	Fill-in forms		
Information sheets	Free	Hypotheses	Total
General			
M	79.7	86.7	83.8
SD	21.7	12.9	16.8
	(*n* = 7)	(*n* = 10)	(*n* = 17)
Domain Specific			
M	77.8	92.6	85.6
SD	11.1	7.8	11.9
	(*n* = 8)	(*n* = 9)	(*n* = 17)
Total			
M	78.7	89.5	84.7
SD	16.3	10.9	14.4
	(*n* = 15)	(*n* = 19)	(*n* = 34)

There is a strong main effect of the type of fill-in form on the total number of forms used. The groups with hypotheses fill-in forms used an average of 5.5 form (SD = 1.5) and the groups with free fill-in forms 3.7 forms (SD = 1.1). The groups with hypotheses forms used a significant higher number of forms than the groups with free forms ($F_{(1,30)} = 15.89$, $p < .01$). The type of information sheet (general vs. domain specific) had no effect on the number of forms used.

The higher activity level for the groups with hypotheses fill-in forms is also found in the number of cells filled in at the forms. The groups with hypotheses forms used an average of 89.5% of the cells (SD = 10.9) and the groups with free forms used 78.7% (SD = 16.3). The groups with hypotheses forms used a significant higher percentage of cells than the groups with free forms ($F_{(1,30)} = 5.31$, $p < .05$).

The groups with hypotheses forms did not make use of the possibility to use free fill-in forms to design extra experiments or to generate and test their own hypotheses. Only one pair (General - Free) used a free form to design an extra experiment. It appears that the hypotheses forms stimulate global activity but does not stimulate free activity. Subjects are passively following the hypotheses offered to them.

3.2.2 *Learning process validity*

At the second level of analysis the statements given by the subjects were assessed on their learning process validity, i.e. if the statement in each cell answered the general description of the cell as it was given on the information sheet. For example: is the hypothesis a description of a relation between two or more variables and/or parameters of the model? Appendix 1 gives the assessment criteria for each of the cells of fill-in forms. Appendix 2 gives a concrete example of the assessment of statements as given by subjects at one fill-in form.

The learning process validity scores were calculated first by counting the number of aspects from the assessment scheme that were included in the statements of the subjects and then relating this score to the maximum the subjects could have scored given the number of cells they used. For example: If one pair of subjects used four PREDICTION cells but only had two valid predictions, they scored 50% on this learning process. Learning process validity is thus a relative score. Groups with hypotheses fill-in forms could not score in the HYPOTHESIS cell and we adjusted the total scores accordingly.
Table 4 presents an overview of the average learning proces validity scores for each of the four experimental groups. The table shows that overall the learning process validity score of the subjects was about 42% (SD = 7.5). The cells VARIABLES & PARAMETERS, EXPERIMENT, and PREDICTION scored around 55 to 57 %. The cells DATA INTERPRETATION and CONCLUSION scored around 25%. The differences in scores between the cells VARIABLES & PARAMETERS, EXPERIMENT, and PREDICTION on the one hand and DATA INTERPRETATION and CONCLUSION on the other are significant (t-tests, $p < .01$).

The data seem to indicate that subjects had difficulties with the learning processes data interpretation and conclusion. We have to remark however that for the cell DATA INTERPRETATION, we included some criteria that were not always necessary for a proper interpretation of the output (see Appendix 1), such as:

- comparison of the output with other graphs or data from previous runs or from other sources (books or lectures);
- comparison of the output with the prediction.

And indeed subjects seldomly noted down these aspects of data interpretations.

Table 4
Mean scores (M) and standard deviations (SD) of the learning process validity for the four experimental groups

	Experimental groups									
	General - Free (*n* = 7)		Domain specific - Free (*n* = 8)		General - Hypotheses (*n* = 10)		Domain specific - Hypotheses (*n* = 9)		Total (*n* = 34)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
VARIABLES & PARAMETERS	57.1	14.8	58.2	13.5	49.5	13.2	58.2	12.0	55.4	13.3
HYPOTHESIS	2.9	7.6	31.3	39.5	-	-	-	-	-	-
EXPERIMENT	56.4	23.6	45.4	19.8	59.6	26.3	65.8	13.2	57.2	21.7
PREDICTION	49.0	41.6	43.8	42.7	67.5	21.7	63.4	24.9	57.0	32.9
DATA INTERPRETATION	24.2	24.4	22.8	7.4	25.0	8.6	25.6	6.9	24.5	12.3
CONCLUSION	11.8	11.8	25.0	0	28.0	4.1	33.1	6.8	25.3	10.0
total learning process validity[3]	44.6	12.7	39.2	5.5	40.0	6.4	44.2	3.8	41.9	7.5

Also the cell CONCLUSION showed a low learning process validity score. This was mainly caused by the fact that subjects hardly generalized their findings to other models. Especially the groups with the free fill-in forms failed to draw conclusions about the validity of the hypotheses.

Table 5 gives an alternative arrangement of the data from Table 4 by displaying the learning process validity scores for each of the main experimental conditions separately. Since we used a nested design each pair of subjects' scores are included in the figures from Table 5 twice, once under the condition 'Information sheets' and once under 'Fill-in forms'.

The only significant influence of the experimental variations on the learning process validity scores can be found for the cell CONCLUSION. Both the factor information sheet ($F_{(1,30)} = 14.09$, $p < .01$) and fill-in form ($F_{(1,30)} = 27.20$, $p < .01$) had a significant effect in this cell. Table 5 shows that the groups with hypotheses forms reached 30.5% (SD = 6.0) of their scores whereas the groups with free forms reached 18.8% (SD = 10.3) of their possible score. Since the former group was provided with learning process valid hypotheses they could possibly more easily review these hypotheses in the conclusion. The groups with domain specific sheets scored 29.3% (SD = 6.4) in the cell CONCLUSION

[3] The total learning process validity is not an average of the scores of the five/six cells. For this score the total absolute score of all cells over all forms is taken and related to the maximum score (given the cells).

and the groups with general information sheets 21.4% (SD = 11.4). We did expect that the domain specific information was helpful for the exploratory learning processes but we did expect an effect on all the processes; not only for conclusions. The group Domain specific - Hypotheses had a high score on three of the six cells, namely EXPERIMENT, DATA INTERPRETATION, and CONCLUSION (see Table 4).

Table 5
Mean scores (M) and standard deviations (SD) of the learning process validity for the information sheet condition and the fill-in form condition

	Information sheets				Fill-in forms			
	Domain specific (*n* = 17)		General (*n* = 17)		Hypotheses (*n* = 19)		Free (*n* = 15)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
VARIABLES & PARAMETERS	58.2	12.3	52.6	14.0	53.6	13.1	57.7	13.6
EXPERIMENT	56.2	19.2	58.3	24.5	62.5	20.8	50.5	21.6
PREDICTION	54.1	34.8	60.0	31.6	65.5	22.7	46.2	40.7
DATA INTERPRETATION	24.3	7.1	24.6	16.3	25.3	7.6	23.4	16.8
CONCLUSION	29.3	6.4	21.4	11.4	30.5	6.0	18.8	10.3
total learning process validity	41.8	5.2	41.9	9.4	42.0	5.6	41.8	9.6

A comparison between the scores of the cell HYPOTHESIS is only possible for the groups that did not receive hypotheses to explore. Therefore we will discuss the results for the HYPOTHESIS cell here separately. The General - Free group only stated one learning process valid hypothesis and this resulted in a learning process validity score of 2.9% (SD = 7.6) (see Table 4). The Domain specific - Free group generated seven hypotheses and this resulted in a score of 31.3% (SD = 39.5) (see Table 4).

Instead of hypotheses subjects noted down other statements in this cell. These statements could be:

- Statements that could be scored under the description of another cell such as VARIABLES & PARAMETERS, PREDICTION and EXPERIMENT;
- Statements that could not be scored at this level (not relevant, not interpretable);
- Statements that expressed a general notion or idea but did not apply to the definition of a hypothesis. These statements can be considered as general explorations of the model. These general model explorations are meant to know the behavior of the system without any experimental manipulations. For example:

  "(...) What is the course deviation over a period of time."

Hypotheses could be stated at a system theory level and at the level of the specific system in the assignment. It showed that subjects did not always make a sensible differentiation between these

levels. Hypotheses were sometimes a combination of both levels.

Five of the hypotheses that were stated by the groups with the free forms were at the system theory level, one was stated at the level of the specific system and two hypotheses were a combination of both levels.

For predictions it was also possible to differentiate between the system theory level and the level of the specific system. For three of the hypotheses stated by the groups with the free forms, the subjects also stated a learning process valid prediction. It appeared that these three predictions were at the same level(s) as the hypotheses.

Furthermore it appeared that subjects sometimes mingled the statements of the HYPOTHESIS and PREDICTION cells. The groups with the free forms gave eight of these mingled statements; four in the cell HYPOTHESIS and four in the cell PREDICTION. The groups with the hypotheses forms made seven of these statements in the cell PREDICTION. An example of a statement that is a mixture of a hypothesis and a prediction is:

"Relative damping of the system is small because the poles are $s=-2*10^{-4}$."

The concept of damping does not imply a choice of experiment, such as determining the poles of the system. This part of the statement applies to a large number of experiments and can be regarded as part of a hypothesis. On the condition side of the statement the poles are mentioned. A statement about the poles implies that a specific choice has been made for a particular experiment. This part of the statement is therefore not a more general statement as would be expected of a hypothesis, but a statement of the output of a specific experiment and as such a part of a prediction.

The data as presented so far assessed statements made in a specific cell (according to the assessment criteria from Appendix 1). However, it happened that subjects noted down a statement in a cell that did not fall under the definition of the learning process of this cell, but that could easily be scored under another cell. For example: experiments were sometimes stated in the cell DATA INTERPRETATION. Tables 6 and 7 (having the same general structures as the Tables 4 and 5) give learning process validity scores if we also take into account statements that were found in the *wrong* cell.

If we use this new way of scoring the total learning process validity of all groups rises from 44.6% (Table 4) to 49.3% (SD = 7.4; Table 6). All cells show a higher score when we score the statements in the present way. The scores in the cells PREDICTION and CONCLUSION do not differ significantly (t-test, $p < .05$) between the two ways of scoring (statements scored in the cell they were placed or statements scored over all cells). The scores in the other four cells did differ (t-test, $p < .01$).The trend in scores, however, is equal; again, the cells VARIABLES & PARAMETERS, EXPERIMENT, and PREDICTION scored around the same mean 64 to 66 %. Also the cells DATA INTERPRETATION and CONCLUSION scored around the same mean 26 to 31%.

Statements that were placed in the wrong cell usually were experiments and data interpretations. Experiments were often placed in the cell DATA INTERPRETATION or HYPOTHESIS. Data interpretations were also placed in other cells, especially if subjects made use of the poles as an experiment. Perhaps subjects did not think of these data as output because these data could be calculated by hand. Statements of the cell CONCLUSION are not often noted down in other cells. Apparently, subjects have a good general idea about this process and this idea does not interfere with other processes.

Statements which are concerned with the system or the relations within the system were also placed in other cells. These statements have to be noted down in the cell VARIABLES & PARAMETERS but were often found in the cell HYPOTHESIS. Especially, the groups with free fill-in forms often noted down the system and the relation within the system in other cells.

Table 6
Mean scores (M) and standard deviations (SD) of the learning process validity for the four experimental groups

	Experimental groups									
	General - Free (*n* = 7)		Domain specific - Free (*n* = 8)		General - Hypotheses (*n* = 10)		Domain specific - Hypotheses (*n* = 9)		Total (*n* = 34)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
VARIABLES & PARAMETERS	75.6	17.5	71.0	19.0	58.8	11.5	62.7	15.3	66.2	16.4
HYPOTHESIS	31.9	19.5	31.3	39.5	-	-	-	-	-	-
EXPERIMENT	65.9	23.1	65.0	22.7	64.3	24.0	70.4	12.1	66.4	20.1
PREDICTION	49.0	41.6	75.0	66.1	67.5	21.7	63.4	24.9	64.4	40.0
DATA INTERPRETATION	28.1	24.9	30.4	12.6	35.3	12.9	29.7	10.1	31.2	15.0
CONCLUSION	11.8	11.8	28.1	8.8	30.8	6.6	33.1	6.8	26.9	11.3
total learning process validity	54.6	11.4	49.5	7.4	45.9	4.3	48.7	4.7	49.3	7.4

An ANOVA showed that for the cell CONCLUSION not only significant main effects from the information sheet ($F_{(1,30)} = 8.66$, $p < .01$) and the fill-in form ($F_{(1,30)} = 16.56$, $p < .01$) can be found (as we have seen in Table 5 before as well), but that, moreover, there is an interaction effect as well ($F_{(1,30)} = 5.71$, $p < .025$). Furthermore, it is notable that the free group has a significant higher score on the cell VARIABLES & PARAMETERS than the hypotheses group ($F_{(1,30)} = 5.30$, $p < .05$). This can be caused by statements about the system and the relations within the system, which were often placed in other cells.

If we compare the hypotheses of the groups with the free forms it becomes clear that General - Free group almost has the same score as the Domain specific - Free group. It is possible that they only had more trouble in placing these statements in the right cell. It is also possible that they found it more difficult to differentiate hypotheses from predictions because in all cases, the hypotheses were stated in the cell PREDICTION.

Table 7
Mean scores (M) and standard deviations (SD) of the learning process validity for the information sheet condition and the fill-in form condition

	Information sheets				Fill-in forms			
	Domain specific (n = 17)		General (n = 17)		Hypotheses (n = 19)		Free (n = 15)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
VARIABLES & PARAMETERS	66.6	17.1	65.7	16.2	60.7	13.2	73.2	17.8
EXPERIMENT	67.9	17.5	65.0	22.9	67.2	19.0	65.4	22.0
PREDICTION	68.8	47.5	59.9	31.6	65.5	22.7	62.9	55.7
DATA INTERPRETATION	30.0	11.0	32.3	18.4	32.6	11.7	29.3	18.6
CONCLUSION	30.8	8.0	22.9	13.0	31.9	6.6	20.5	13.0
total learning process validity	49.1	5.9	49.5	8.9	47.2	4.6	51.9	9.5

3.2.3 *Domain correctness*

At the third level of analysis we assessed the statements on their domain correctness. We only analyzed the aspects of the statements that were valid at the previous learning process validity level (see Appendix 1). For each aspect that was scored at the previous level we determined if this aspect of the statement was also correct in domain terms. Our general idea was that the following analysis levels, consistency and overall strategy, could not be performed if the statements were domain incorrect. If a subject made a mistake in one of the cells which caused mistakes in other cells, then the mistake was only scored once. Of course, subjects could state 'false' hypotheses and, therefore, the HYPOTHESIS cell is analyzed separately from the other data. If subjects have stated a 'false' hypothesis, then the prediction was not scored incorrect if it was in line with this hypothesis. Actually this only occurred once.

The domain correctness scores were relative scores (percentages) in the sense that they were related to the scores of the previous level (learning process validity). For example, if a pair of subjects had two valid experiments, but one of them was incorrect, they scored 50% on domain correctness.

Table 8 shows that overall there is a high score on domain correctness (80.4%, SD = 12.7) of the aspects of the learning process valid statements. So, once the subjects made statements that were valid as an exploratory learning process, these statements were mostly correct at a domain level. The trend that we saw at the previous level, in which the cells DATA INTERPRETATION and CONCLUSION scored lower than the other cells, is not seen at the domain correctness level. All the cells score percentages from 70 to 85%.

Again we can see significant effects from the experimental condition Fill-in form on the scores in cell

CONCLUSION ($F_{(1,30)} = 11.52$, $p < .01$). Table 9 (presenting a similar arrangement of data as the Tables 5 and 7) shows that the groups with hypotheses forms scored 91.1% (SD = 14.3) of their scores whereas the groups with free forms scored 81.1% (SD = 38.7). For the total scores on the third analysis level there is a similar effect ($F_{(1,30)} = 6.10$, $p < .025$).

We have to remark that one of the hypotheses we gave the subjects was quite easy. This may have influenced this score. Furthermore, we had offered the nine hypotheses in a certain order. The more complex hypotheses were at the end of this order. Since the mean of the number of forms of the hypotheses groups is 5.5, it is possible that some of the subjects in these groups did not work on the more complex hypotheses.

Contrary to what might have been expected, the factor of the information sheet did not show an effect. The groups with domain specific information did not score higher on this level. As with most of the scores on the previous levels, the group Domain specific - Hypotheses had the highest score on all the cells.

Table 8

Mean scores (M) and standard deviations (SD) of the domain correctness score (per cell and total) for the four experimental groups

	Experimental groups									
	General - Free (n = 7)		Domain specific - Free (n = 8)		General - Hypotheses (n = 10)		Domain specific - Hypotheses (n = 9)		Total (n = 34)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
VARIABLES & PARAMETERS	67.7	24.8	77.5	24.8	77.1	21.3	92.8	5.1	79.4	21.3
EXPERIMENT	82.8	19.3	73.9	30.7	89.2	13.7	92.6	11.9	85.2	20.1
PREDICTION	71.4	48.8	50.0	46.3	65.8	39.0	92.6	14.7	70.3	40.0
DATA INTERPRETATION	55.1	41.0	81.6	23.9	68.9	21.3	86.6	17.0	73.7	27.6
CONCLUSION	45.2	45.9	75.0	26.7	87.2	18.1	95.4	7.4	77.9	31.2
total domain correctness[4]	76.0	13.3	74.5	14.9	78.8	10.7	90.8	5.1	80.4	12.7

[4] The total domain correctness is not an average of the relative scores of the five cells. For this score the total absolute score over all cells and forms is taken and related to the scores of the previous level.

Table 9
Mean scores (M) and standard deviations (SD) of the domain correctness score (per cell and total) for the information sheet condition and the fill-in form condition

	Information sheets				Fill-in forms			
	Domain specific (n = 17)		General (n = 17)		Hypotheses (n = 19)		Free (n = 15)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
VARIABLES & PARAMETERS	85.6	18.6	73.2	22.6	84.5	17.4	73.0	24.4
EXPERIMENT	83.8	24.0	86.6	16.0	90.8	12.7	78.1	25.5
PREDICTION	72.5	39.1	68.1	41.9	78.5	32.3	60.0	47.1
DATA INTERPRETATION	84.2	20.0	63.2	30.6	77.3	20.9	69.2	34.5
CONCLUSION	85.8	21.2	69.9	37.8	91.1	14.3	61.1	38.7
total domain correctness	83.1	13.4	77.6	11.6	84.5	10.4	75.2	13.7

Since subjects stated little hypotheses, we looked at the absolute numbers for this cell. The hypothesis that was generated by the General - Free group was domain correct. The seven hypotheses of the Domain specific - Free group that were learning process valid, were almost all correct; just one was domain incorrect. Three other hypotheses could be not be verified.

In general, mistakes were caused by a wrong apprehension of some of the concepts, especially if it involved the frequency domain. Furthermore, one typical domain mistake was made by a number of the subjects. Subjects quite often lacked insight in control theory and this resulted in the creation of a wrong model of the system. We like to explain this mistake in more detail:

> The system presented to the students in the assignment is a cargo freighter. The aim of the assignment is to design an automatic pilot for the ship. This automatic pilot determines the course of the ship and adjusts its rudder in order to keep it on course. The ship is perceived as a system. The output signal of the system is the course of the ship. The input signal consists of forces applied the ship by the rudder and wind, waves and currents. Only the forces applied to the rudder can be controlled. A common mistake made by the students as they model the situation is that they do not realize that the forces by wind, waves and currents are uncontrollable. Therefore, the model they create and the results of the simulation will be incorrect.

3.2.4 *Consistency*

At the fourth level we assessed the relations between the statements in different cells on one fill-in form. The following relations were evaluated:

- HYPOTHESIS - EXPERIMENT

- EXPERIMENT - PREDICTION
- EXPERIMENT - DATA INTERPRETATION
- PREDICTION - DATA INTERPRETATION
- DATA INTERPRETATION - CONCLUSION

We only analyzed statements that were valid at Level 2 (learning process validity) and domain correct. For learning process validity we made a small adjustment. Not all the aspects of Appendix 1 had to be scored positively. For each cell we determined the aspects that were absolutely necessary. In Appendix 1 they are marked with a '*'.

For the cell HYPOTHESIS we also analyzed hypotheses which were scored domain incorrect because they could not be verified. This is the only exception to our idea of the analysis level as a 'sieve'. It did not influence the results because the relation HYPOTHESIS - EXPERIMENT was analyzed separately from the rest of the data.

Again, consistency scores were relative scores (percentages). The scores were related to the maximum possible score a pair of subjects could get. For example: If subjects had given all the necessary aspects in all the cells and made no domain mistakes, they could score all five relations. If half of these relations were correct they would receive a score of 50%.

Table 10 shows that 90.1% (SD = 24.1) of the relations were correct. It appears that learning process valid and domain correct statements result very often in correct relations.

Table 10
Mean scores (M) and standard deviations (SD) of the consistency score

	Fill-in forms		
Information sheets	Free	Hypotheses	Total
General			
M	80.1	86.5	83.9
SD	36.7	31.3	32.6
	(*n* = 7)	(*n* = 10)	(*n* = 17)
Domain Specific			
M	95.8	96.6	96.3
SD	8.9	5.2	7.0
	(*n* = 8)	(*n* = 9)	(*n* = 17)
Total			
M	88.5	91.3	90.1
SD	26.1	23.0	24.1
	(*n* = 15)	(*n* = 19)	(*n* = 34)

At the consistency analysis level we did not find any effects of the fill-in forms or information sheets ($p < .05$). At the previous three levels we already saw that the group Domain specific - Hypotheses had high scores. This is the same for the present analysis level.

For the groups with hypotheses forms we could look at the relation between HYPOTHESIS - EXPERIMENT. Seventy-one of these relations could be analyzed and we found that only five of these

relations were incorrect. It appears that if hypotheses are provided, then the design of a matching experiment is not so difficult. However, the relation at this general level does not say anything about the appropriateness of the experiment. Is the experiment a critical one, will it verify or falsify the hypothesis, etc.

3.2.5 *Overall strategy*

At the final analysis level we wanted to assess the development of the successive hypotheses and the relation between a hypothesis and the conclusion on the previous form. As we mentioned in the preceding section, very few statements passed the 'sieve'. Therefore, we only discuss some examples.

Only a few hypotheses were generated and they usually did not followed one another. In three cases the hypotheses were successive. One time the transition was from a hypothesis about the system *without* a control law to a hypothesis about the system *with* a control law. The other two cases were concerned with hypotheses about the system with different control laws; going from a simple to a more complex control law.

A relation between a hypothesis and a conclusion on the previous form occurred four times. Two of them involved the transition from a conclusion about the system *without* a control law to a hypothesis about the system *with* a control law. The other two transitions involved systems with two different control laws; one simple law with one feedback amplification and one more complex control law. we like to illustrate this with the following example:

CONCLUSION
"A proportional control law can not prevent oscillation. However, in the long run it will bring the ship on course."

HYPOTHESIS (next form)
"A PI control law (proportional and integrating) will do the job. Disruptions of the system occur mainly in the lower frequency range".

The link between the conclusion and the following hypothesis is primarily the observation that you need more than a proportional control law. The subjects just added one element to the control law.

3.2.6 *Post-test*

At the end of the lab session, all subjects made a post-test which consisted of seven multiple choice questions. These questions tested qualitative insight in the domain. The subjects could get a score of 7.0. The test was not extensive because of the limited time available.

The mean of the score of all subjects was 4.8 (SD = 1.3). Three of the four experimental groups scored 4.7 or 4.8. The only exception was the group Domain specific - Hypotheses; they scored 4.3 (SD = 1.6). The control group had a higher score of 5.2 (SD = 1.1). The only significant difference was between these two groups ($t = 2.33$, $p < .05$).

At first sight, the results of the post-test would imply that the traditional computer lab of the control group showed the highest benefit and that the experimental group which received the most support showed the lowest benefit. Actually, this was also found in other studies in which a simulation with instructional support was offered (e.g. de Jong, de Hoog & de Vries, 1991). Some other studies show

that the amount and type of support does not always have the effect that is expected. For example in Shute's (1990) study it was found that subjects with high inductive reasoning skills profited from a deductive instructional environment.

4 Conclusions

In the studies that we presented here we tried to gain insight in exploratory learning processes. We also evaluated the impact of instructional support measures on these processes.

In charting the exploratory study process (Study 1) we identified a total number of 22 detailed learning processes. Two main categories of learning processes that we could distinguish were *transformative processes* and *regulative processes*. Through analyzing thinking aloud protocols of students working with a computer simulation we found that students were not willing to exploit learning processes that are characteristic for exploratory learning (such as hypothesis generation etc.). Furthermore, we found that regulative processes play a fairly important role in exploratory learning certainly if we compare this to non-exploratory learning as is learning from text (Ferguson-Hessler & de Jong, 1990).

In Study 1 we also found that exploratory learning poses problems to students. Students need quite a lot of extra *domain information*. This is in line with the general notion that for fruitful exploratory learning a sufficient level of prior knowledge is necessary (e.g. Flick, 1990).

Finally, the results of Study 1 show that exploratory learning processes are not stimulated by giving instructional support in the form of directive hints.

The fact that students were reluctant to show exploratory behavior could have two reasons: they simply do not want to work in an exploratory way or they have *problems with the exploratory processes* as such. Results from the second study give evidence in support of the second reason.

Subjects in Study 2 received information sheets with general and/or domain specific information on the exploratory learning processes. The other experimental condition consisted of fill-in forms (sheets with six separate areas for six different exploratory learning processes) with or without hypotheses already given. The statements that the students noted down at the fill-in forms were analyzed in a step wise order. A number of levels of analysis were used and each level worked as a 'sieve' for the next level. First we looked at the *global activity level* of the subjects, second the *learning process validity* of each statement made was assessed, third the *domain correctness* of each (learning process valid) statement was determined, fourth we looked at the *consistency*, the relations between different (learning process valid and domain correct) statements. Finally, at the fifth level the *overall strategy* of the students was assessed. At each level subjects gained scores that expressed a percentage of correctness relative to some maximum score. The overall (over all experimental conditions and all subjects) average scores for each level are depicted in Table 11.

Table 11 depicts at each level percentages (with of course a maximum score of 100%) and this makes that we are able to compare the different levels. The table clearly shows that the bottle neck in scientific discovery learning is the valid performance of exploratory learning processes. Students are quite active, and once they have made a statement that is valid in terms of a specific learning process they do not seem to have trouble with the domain itself or with relating different statements to each other.

More specifically we see that *generation of hypotheses* is a difficult task. This learning process scored low; 15 pairs of subjects stated only a total of eight hypotheses. Designing experiments and predicting outcomes, however, are performed better. Subjects had considerable trouble with an adequate performance on data interpretation and drawing conclusions. Another problem for the students was the difference between a hypothesis and a prediction. These two processes were

frequently mixed up.

Both Study 1 and 2 show that data interpretations were mostly done at a rather shallow level. In Study 1 subjects usually made local interpretations; in Study 2 they scored low on this learning process. For the process of evaluating, both studies show that *generalizations* were hardly made.

The need for domain specific information is not so strongly found in Study 2; domain correctness has a high score and the domain specific information sheets showed little effect. It is likely that domain information should be strongly related to specific exploratory learning processes.

Table 11
Mean scores (M) and standard deviation (SD) of the total scores (over all conditions, subjects and different learning processes) on four analysis levels[5]

	Total score	
Analysis level	*M*	*SD*
Activity level (cells)	84.7	14.4
Learning process validity	41.9	7.5
Domain correctness	80.4	12.7
Consistency	90.1	24.1

A second major aspect of our studies was the *impact* of the *different instructional support measures* that we designed. Compared to Study 1, the assignment in Study 2 required more free activity and the instructional support consisted of a combination of directive and non-directive support. Although a direct comparison to the results of Study 1 is impossible, we feel that the subjects in Study 2 show a higher activity on the exploratory learning processes. This impression is supported by the global activity level as depicted in Table 11. Within Study 2, there was a consistent trend in the data that the experimental group with the highest level of instructional support (Domain specific - Hypotheses) scored high on all analysis levels that we used.

In Study 2 we found a number of effects of the instructional measurements in this study. At the global activity level it showed that the groups with hypotheses fill-in forms were more active in the number of forms used and in the relative number of cells filled in. However, having hypotheses available did not stimulate the subjects to create additional hypotheses themselves. The groups that had to generate their own hypotheses were less active. Apparently, generation of hypotheses is a time consuming activity.

The only effects of variations in information sheets and fill-in forms were found for the cell CONCLUSION. The groups that were provided with hypotheses more frequently stated correct conclusions and this also was true for the groups with the domain-specific information sheets.

An important question in studies like this one is of course: what is *the effect on performance measures*. Due to practical reasons the test that we could give in Study 2 was a very limited one. Results, however, did certainly not show an advantage of the experimental groups over the control group. Results of recent studies (van Joolingen & de Jong, 1991; de Jong, de Hoog & de Vries,

[5] At the fifth level of analysis 'overall strategy' we especially wanted to analyze the development of successive hypotheses and the relation between a hypothesis and the conclusion on a previous fill-in form. At the learning process validity level it already showed that subjects stated little hypotheses. So, we did not have sufficient data for analysis at the fifth level.

1991) showed that students were less active if they received a form of instructional support that offered them a lot of information. These studies conclude that students were probably restrained by the amount of information. The students had to read and understand all the information offered to them and were therefore less active. These results are not found in our study. The group which received support with the largest amount of information, the group with domain specific sheets and hypotheses forms, scored high on both global activity measures. Also, the groups with domain specific sheets showed no significant differences with the groups which received general information sheets.

Furthermore, the complexity of the task will probably also influence global activity. Subjects might be less active, not only because they had to read a lot of information, but also because they were involved in a cognitive complex task. The task of the groups with the free forms was more demanding; they had to formulate the hypotheses themselves. So, even if they had less information to read (e.g. the General - Free group), the complexity of their task might have restricted their global activity. Our results show that the groups with free forms were indeed less active. Also, the group with the highest level of support (Domain Specific - Hypotheses) had the lowest score on the post test. As we look at the data at the five analysis levels, the Domain specific - Hypotheses group always have high scores. So, this group was more active and made better use of the exploratory learning processes. It is possible that they were more involved in the new aspects of exploratory learning. We should considered a different (pre- and) post-test which would also test the subjects' scientific inquiry abilities.

The findings that we reported here should be seen in the *context of our research*. We try to perform experiments within regular curricula. This means that both experimental rigor and possibilities for implementing experimental conditions are endangered. An example of this is, that we, for reasons of time, could only administer a very limited post-test. The great advantage is that research is done with subjects in a situation for which the conclusions are highly relevant. For example, the experimental conditions in Studies 1 and 2 are still very close to what is practically or politically feasible in the instructional context.

A main consequence of our approach is that the impact of the instructional measures is necessarily a small one compared to the rest of the instruction our subjects receive in the domain (topic) involved in these studies. This also means that students that participate in the experiment hardly have time to get adjusted to a for them brand new way of learning and instruction. Despite this the studies reported have revealed some areas that pose specific problems to students. Also, the instructional measures that we implemented have shown some effects. It is, however, likely that students need to work in this way for a longer period of time, before more significant results will appear.

REFERENCES

Ausubel, D.P., Novak, J.D., & Hanesian, H. (1978). *Educational psychology: A cognitive view.* New York: Holt, Rinehart & Winston.

van Berkum, J.J.A., & de Jong, T. (1991). Instructional environments for simulations. *Education & Computing, 6*, 305-358.

Böcker, H.D., Herczeg, J., & Herczeg, M. (1989). ELAB - An electronics laboratory. In D. Bierman, J. Breuker & J. Sandberg (Eds.), *Proceedings of the 4th international conference on AI & Education* (pp. 15-25), Amsterdam: IOS.

Cronbach, L.J. (1977). *Educational Psychology*. New York: Harcourt Brave Jovanovich.

Ferguson-Hessler, M.G.M., & de Jong, T. (1990). Studying physics text; differences in study processes between good and poor performers. *Cognition and Instruction, 7*, 41-54.

Finegold, M., & Gorsky, P. (1989). Learning about forces: simulating the outcomes of pupil's misconceptions. *Instructional Science, 17*, 251-261.

Flick, L.B. (1990). Interaction of intuitive physics with computer simulated physics. *Journal of Research in Science Teaching, 27*, 219-231.

Hartley, J.R. (1988). Learning from computer based learning in science. *Studies in Science Education, 15*, 55-76.

Hille, R. (1980). Denk- und lernpsychologische aspekte von unterrichtseinheiten zur computer-simulation und modellbildung. In Simon, H. (Ed.), *Computer-simulation und modellbildung im unterricht.* Munchen/Wien: R.

Oldenbourg Verlag.

de Hoog, R., de Jong, T., & de Vries, F. (1991a). Interfaces for instructional use of simulations. *Education & Computing, 6*, 359-385.

de Jong, T. (1990). *Ontwikkelingen in Computer Ondersteund Onderwijs in het Nederlands Hoger Onderwijs in de jaren 1988/1989.* [Developments in computer assisted instruction in the Netherlands for the years 1988/1989]. Eindhoven University of Technology: OCTO rapport 90/01.

de Jong, T. (Ed.) (1991a). Computer simulations in an instructional context. *Education & Computing* (special issue).

de Jong, T. (1991b). Learning and instruction with computer simulations. *Education & Computing, 6*, 217-229.

de Jong, T., de Hoog, R., de Vries, F. (1991). SUPER-MIDAS, a computer simulation for learning about decision support theory. University of Amsterdam (Department of Social Science Informatics): SAFE/SIM/WPII/UvA-rep/SUPER-MIDAS.

de Jong, T., & Njoo, M. (1991). *Learning and Instruction with computer simulations: learning processes involved.* To appear in the proceedings of NATO ARW, Leuven september 1990.

van Joolingen, W.R., & de Jong, T. (1991a). Characteristics of simulations for instructional settings. *Education & Computing, 6*, 241-262.

Van Joolingen, W.R., & De Jong, T. (1991b). A prototype scratchpad for hypothesis formation and experiment design. In: Hijne (ed.), *Prototype/mock-up of an Integrated Simulation-based Learning Environment, SIMULATE deliverable SIM/23.* Zaandam: Courseware Europe bv.

Kirby, J.R. (1984). Strategies and processes. In J.R. Kirby (Ed.), *Cognitive strategies and educational performance* (pp. 3-12). New York: Academic Press.

Klahr, D., & Dunbar, K., (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1-48.

Kearsley, G. (1985). Embedded training: The new look of computer-based instruction. *Machine-mediated learning, 1*, 3, 279-296.

Lavoie, D.R., & Good, R. (1988). The nature and use of prediction skills in a biological computer simulation. *Journal of Research in Science Teaching, 25*, 5, 335-360.

Lesgold (1990). Tying development of intelligent tutors to research on theories of learning. In: H. Mandl, E. de Corte, S.N. Bennett, & H.F. Friedrich (Eds.), *Learning & Instruction; European research in an international context* (vol. 2.1). Oxford: Pergamon.

Lodewijks, H.G.L.C. (1985). Wat is onderzoek van onderwijsleerprocessen? *Jaarverslag jaarboek 1984.* 's Gravenhage: SVO.

Michalski, R. (1987). Learning strategies and automated knowledge acquisition. In G.L. Bradshaw, P. Langley, R.S. Michalski, S. Ohlsson, L.A. Rendell, H.A. Simon, & J.G. Wolff (Eds.), *Computational models of learning* (pp. 1-19). Bern: Springer-Verlag.

Njoo, M., & de Jong, T. (1991). Learning processes of students working with a computer simulation in mechanical engineering. To appear in *Proceedings of the EARLI third European conference for research on learning and instruction.* Oxford: Pergamon.

Rivers, R.H., & Vockell, E. (1987). Computer simulations to stimulate scientific problemsolving. *Journal of Research in Science Teaching, 24*, 403-415.

Reimann, P. (1989). Modeling scientific discovery learning processes with adaptive production systems. In D. Bierman, J. Breuker & J. Sandberg (Eds.), *Artificial intelligence and education; synthesis and reflection. Proceedings of the 4th International Conference on AI and Education* (pp. 218-227). Amsterdam: IOS.

Shute, V.J. (1990). A comparison of inductive and deductive learning environments: Which is better for whom an why? *Paper presented at the 1990 AERA conference.* Boston (USA).

Shute, V.J., & Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments, 1*, 51-77.

Swanson, J.H. (1990). The effectiveness of tutorial strategies: an experimental evaluation. *Paper presented at the 1990 AERA conference.* Boston (USA).

Towne, D.M., Munro, A., Pizzini, Q.A., Surmon, D.S., Coller, L.D., & Wogulis, J.L. (1990). Model-building tools for simulation-based training. *Interactive Learning Environments, 1*, 33-50.

Veenman, M.V.J., & Elshout, J.J. (1990). De meerwaarde van een goede probleemaanpak. [The surplus value of a proper problem approach]. *Tijdschrift voor Onderwijsresearch, 15*, 337-347.

White, B.Y., & Frederiksen, J.R. (1987). Qualitative models and intelligent learning environments. In: W. Lawler & M. Yazdani (Eds.) *Artificial intelligence and education* (pp. 281-305). Norwood: Ablex.

White, B.Y., & Frederiksen, J.R. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence, 42*, 99-157.

APPENDIX 1 - Assessment scheme for determining learning process validity.

Criteria for assessing statements in the different cells.

VARIABLES&PARAMETERS
- Variables/parameters*[6]
- Differentiation between variables and parameters
- Differentiation between independent and dependent variables
- Relations between variables and/or parameters
- System

HYPOTHESIS
- Relation between two or more variables and/or parameters*

  At the level of system theory, this has to be a statement about the properties of a class of systems.
  At the level of a specific system, this has to be a statement about the behavior of the system in time and for any input signal.

EXPERIMENT
- Input variables/parameters*
- Output variables/parameters*
- (Range of) values of input variables/parameters

PREDICTION
- Results of the experiment (qualitative/quantitative)*

  At the level of system theory, this is a statement about the output, given a system (specified in the experiment).
  At the level of a specific system, this is a statement about the output, given a certain input signal.

DATA INTERPRETATION
- Output (graphs/data)*
- Discussion of characteristics of the output*
- Discussion of the properties of the (class of) system(s)*
- Comparison with the prediction
- Comparison with other graphs/data
- Discussion about the experiment as a test for the hypothesis

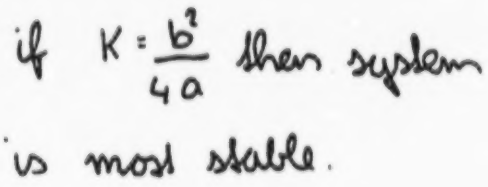
CONCLUSION
- Validity of the hypothesis*
- Discussion of the model*
- Generalization to other models
- Generalization to control theory

[6] * = Aspects that absolutely had to be present for analysis of the consistency level (See section 3.2.4).

APPENDIX 2 - Example of data analysis at Level 2: Learning process validity

We will illustrate the assessment of learning process validity by discussing the statements of one pair of students. This pair was a member of the group that received domain specific information and had free fill-in forms. The form shown was the third form that they had filled in.

In the cell 'Hypotheses' these students noted down:

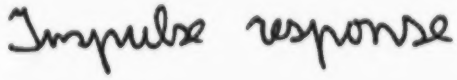


This means that they state a optimal choice for the feedback amplification (K) with regard to the stability of the system.

At the previous form students had already made the choice for a specific type of control law (a *proportional* control law) and now they make a choice for the value of the parameter in this law.

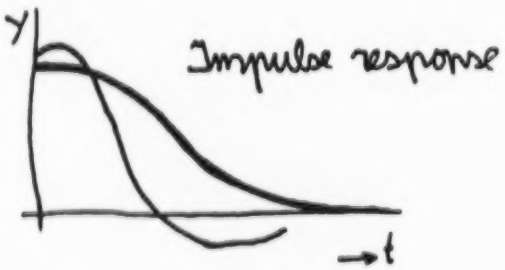
This statement received the full score for stating a hypothesis. Both dependent (stability) and independent variables (K) and a relation (optimal level) between them was indicated.

In the cell 'Experiment' the students only noted down:



This means that they indicated the output that they wanted to look at. They did not specify which variables (in the concrete situation) to vary, the values to give to the variables to be manipulated, and about time scale at which to evaluate (the impulse response). As a consequence they received only one out of three possible credit points for the cell 'Experiment'.

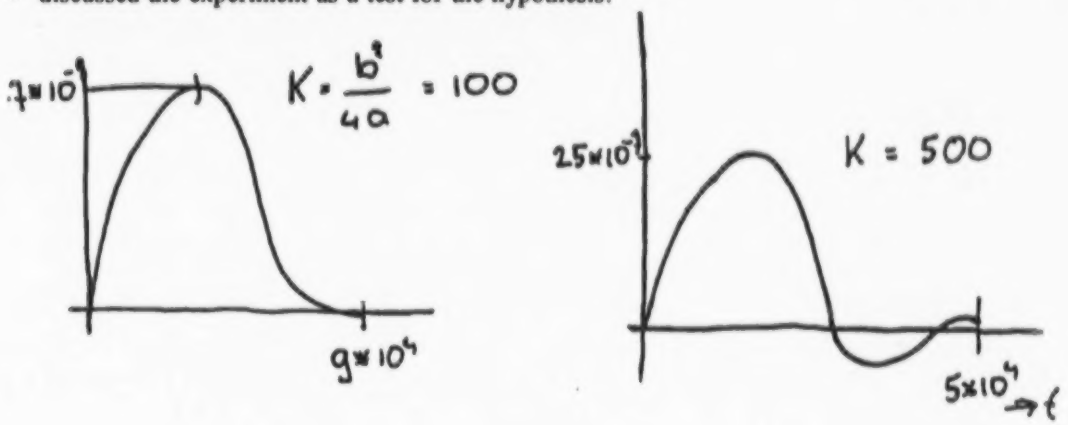
In the 'Prediction' cell the students correctly stated a prediction, indicating the expected results of an experiment. The have done this in a graphical way:



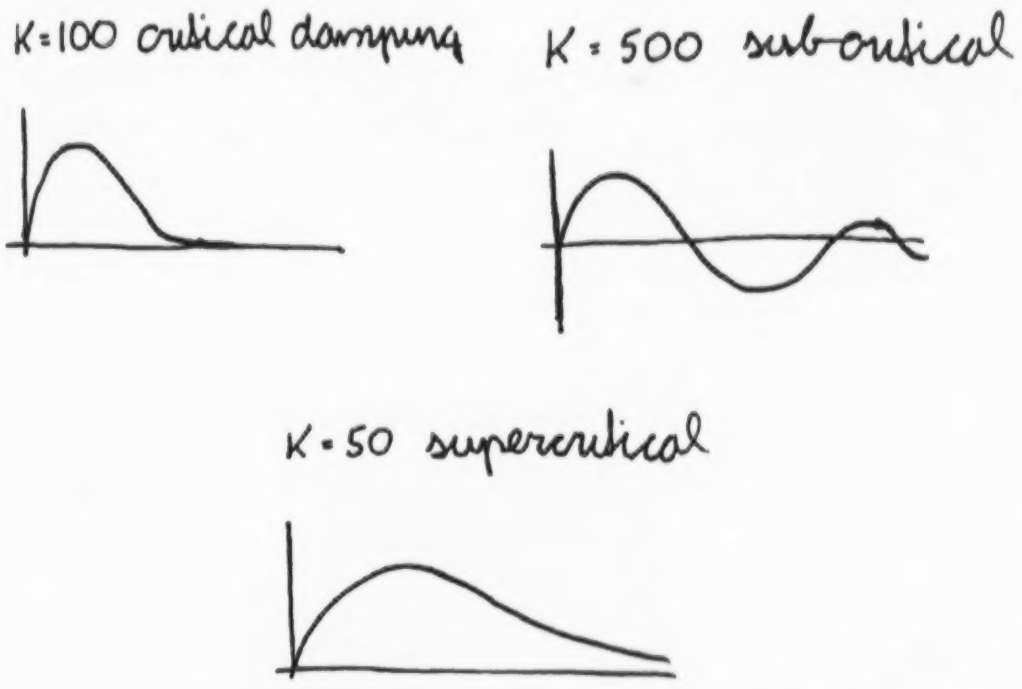
For 'data interpretation' we used six criteria in our assessment. As is shown below the students indicated at their form only one of these: they noted down the (graphical) output of the simulation run. They have chosen two values for K. Next to the one they have indicated in the hypothesis, they

have also taken another value for K (K = 500) to show that with this K it takes more time to return to the initial position. The criteria they have 'missed' are that they have not:

- discussed the characteristics of the output
- discussed important properties of the (class) of system(s)
- made a comparison to the prediction
- made a comparison to other experiments, graphs, or data
- discussed the experiment as a test for the hypothesis.



Finally, in the conclusion section we find the following:



What we see here is that they give some data interpretation. They indicate important characteristics of the system by stating for which K there is *critical damping*, *sub critical damping*, and *super critical damping*. Here they conclude that the hypothesis that was stated is incorrect. This is, by the way, a correct conclusion since there is a whole range of K's for which stability is maximal.